

Iterative Collaborative Representation based Classification for Face Recognition

Wei Huang^{*1}, Xiaohui Wang¹, Yinghui Zhu¹, Jianzhong Li²

Department of Computer Science and Engineering, Department of mathematics and statistics, Hanshan Normal University, Chaozhou, Guangdong Province, China

^{*1}hw@hstc.edu.cn.edu; ²wangxh@hstc.edu.cn; ³yhzhu@hstc.edu.cn; ⁴henry_stu@163.com

Abstract

Collaborative representation based classification (CRC) has received much attention in the field of pattern recognition. CRC uses a simple way to code a testing sample as a linear combination of all the training samples and classifies the testing sample into the class with the minimum representation error. But the original algorithm of CRC suffers from the following problem. It must perform the matrix inverse operation, which may cause unstable numerical computation. With this paper, in order to overcome the above problem of the original algorithm of CRC, we propose an iterative collaborative representation based classification (ICRC) algorithm. The experimental results on face recognition show that ICRC not only outperforms CRC but also is able to obtain a much higher accuracy than LRC.

Keywords

Face Recognition; Collaborative Representation; Iterative; Classifier; CRC

Introduction

Face recognition has received much attention in the last few decades [1,2,3]. Though many methods have been proposed in the past, face recognition is still faced with great challenges, which motivates researchers to devote their efforts to this field [4-8].

Feature extraction and classification are two important aspects of face recognition. Most of previous feature extraction methods could be categorized into two categories namely unsupervised and supervised methods. Principle component analysis (PCA) [9] and independent component analysis (ICA) [10] are famous unsupervised feature extraction methods. Typical examples of supervised feature extraction methods include linear discrimination analysis (LDA) [11], uncorrelated linear analysis (ULDA) [12] and so on.

Many classification methods can be used in face recognition [13, 14, 15]. The nearest neighbor classifier may be the most widely used method owing to its simplicity and effectiveness. Many variants of the nearest neighbor classifier including the local mean based nearest neighbor classifier [16], adaptive nearest neighbor classifier [17], K nearest neighbor classifier [18] and center-based nearest neighbor classifier [19] have also been used in face recognition.

Recently, linear regression classification (LRC) [20] has been proposed for face recognition and obtained promising performance. LRC is based on the assumption that face images from the same class lie on a linear subspace. It seems that LRC is directly related with the nearest space method. LRC can be easily implemented by solving linear equations. Wright et al. proposed a sparse representation based classification (SRC) [21] and applied it for face recognition successfully. SRC codes a testing sample by sparse linear combination of all training samples and then classifies the testing sample into the class with the minimum representation error. SRC is different from LRC as follows. First, SRC exploits a sparse linear combination of all training samples to represent the testing sample whereas LRC just uses a linear combination of the training samples of a class to express the testing sample. Second, SRC is based on the ℓ_1 regularization but LRC is not. The ℓ_1 regularization means that SRC should seek its solution under the condition that the ℓ_1 norm of the solution vector is the minimum. This implies that SRC does not have a simple closed-form solution though what it addresses is a convex problem. Actually, SRC usually designs an iterative algorithm for calculating its solution. Differing from SRC, LRC has no similar constraint when it seeks its solution. For this reason, we can refer to SRC and LRC as ℓ_1 regularization based representation

classification method and simple representation classification method, respectively. It is clear CRC is also a simple representation classification method.

Before CRC is proposed, the ℓ_1 regularization is taken as the most significant core of SRC. However, Zhang et al. analyzed SRC deeply and pointed out that collaborative representation was more important than the ℓ_1 regularization [22]. So they proposed collaborative representation based classification (CRC). Extensive experiments demonstrated the notable performance of CRC. Variants of CRC such as the method in [23] are also proposed for face recognition. The method in [23] can be viewed as an improvement to CRC. Because it exploits only the training samples of a small number of classes to represent the testing sample, superficially it obtains sparse representation of the testing sample via training samples.

CRC must calculate the matrix inverse operation, which may cause unstable numerical computation. Also, face recognition is the typical small size sample problem [24]. The dimension of the training sample is much more than the number of the training samples. So the testing sample can not be accurately represented by the training samples. These are the major factors which affects the performance of CRC.

In this paper, we propose an iterative algorithm for CRC. In this algorithm, the matrix inverse operation is avoided, so the numerical stability caused by the matrix inverse operation can be eliminated. The rest of the paper is organized as follows: in Section 2, we briefly review LRC and CRC algorithms. In Section 3 iterative CRC is described. In Section 4, several experiments are present to demonstrate the improvement of proposed algorithm. Conclusions are summarized in Section 5.

Related works

Linear Regression Classification

Linear regression classification is based the assumption that a specific class lie on a linear subspace. Under this assumption, the testing sample is represented as a linear combination of class-specific training samples by using the least square method. Finally, the testing sample is signed as the class with the minimum reconstruction error.

Assume we have c different individuals. Let x_i denotes the i^{th} ($i \in \{1, 2, \dots, n_p\}$) face images of the p^{th} subject, where d is the dimensionality of the face image vectors and n_p is the number for the p^{th} subject. X_p is a class-specific model generated by stacking the n -dimensional image vectors, $X_p = [x_1^p, x_2^p, \dots, x_{n_p}^p] \in R^{d \times n_p}$.

Given a testing sample $y \in R^d$, LRC solves the following linear regression problem to get the representation coefficients $\hat{\alpha}_p \in R^{n_p \times 1}$:

$$\hat{\alpha}_p = \arg \min \|y - X_p \alpha_p\|_2 \quad \forall p \in (1, 2, \dots, c) \quad (1)$$

where $\|\cdot\|_2$ stands for the ℓ_2 norm.

$\hat{\alpha}_p$ can be estimated by suing least-squares estimation:

$$\hat{\alpha}_p = (X_p^T X_p)^{-1} X_p^T y \quad (2)$$

The testing sample can be reconstructed from Eq.(3)

$$\hat{y}_p = X_p \hat{\alpha}_p = X_p (X_p^T X_p)^{-1} X_p^T y \quad (3)$$

The distance measure between the testing sample y and reconstructed sample $\hat{y}_p$ ($p = 1, 2, \dots, c$) can be calculate from Eq.(4)

$$dp(y) = \|y - \hat{y}_p\|^2 \quad i = 1, 2, \dots, c \quad (4)$$

The identity of testing sample y is retrieved by the minimum reconstruction error as follows:

$$identity(y) = \arg \min_p dp(y) \quad (5)$$

Collaborative Representation Classification

Zhang et. al in reference [22] pointed that collaborative representation was the real power of SRC. As a typical small-sample-size problem, X_p is always not enough to represent the testing sample y well even when y is from class p . So collaborative representation classification is proposed, which seeks to represent y collaborative by using all the training samples from different classes. Assume $X = [X_1, X_2, \dots, X_p, \dots, X_C]$, $X_p \in R^{d \times n_p}$, n_p is the number for the p^{th} subject. X includes all the training from all the classes. $\Psi = [\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_c]$ ($\alpha_k \in R^{1 \times n_p}$) is the coefficient of X . So the coefficient vector Ψ can be estimated by:

$$\Psi = (X^T X + \lambda I)^{-1} X^T y \quad (6)$$

where λ is a positive small constant and we set $\lambda = 0.01$ in this paper. Eq.6 shows that CRC uses all the training samples to represent testing samples. The representation residual e_p can be calculated as follows:

$$e_k = \|y - \alpha_k X_k\|^2 \quad k = 1, 2, \dots, c \quad (7)$$

The rule in favor of the class with minimum distance can be calculated by:

$$identity(y) = \arg \min_k e_k \quad (8)$$

Iterative CRC

In this section, we introduce the iterative CRC algorithm for face recognition. From the description in section 2, we can find some similarities of LRC and CRC. Firstly, LRC and SRC all play a role as a ℓ_2 norm constraint on represented coefficients. Secondly, the classification criteria of LRC and CRC are based on the minimum representation residual. The difference between LRC and CRC only lies on the collaborative representation. In LRC, a testing sample is represented by training samples from a specific class. But in the original CRC algorithm, a testing sample is represented by all the training samples from all the classes. In face recognition, the number of training samples is far less than the dimension of the training samples. Thus, both LRC and CRC usually suffer from the problem that the testing sample can not entirely be represented by linear combination of the used training samples. Moreover, LRC and CRC have to perform the matrix inverse operation. It is well known that inverse operation may cause unstable numerical computation.

The iterative method has been introduced to the field of sparse representation and promising results can be obtained. For example, Michal [25] proposed an iterative method, which is called K-SVD, to generate sparse coding of samples based on the dictionary and to update the dictionary atoms. Xu et. al proposed a two-phase test sample sparse representation method for face recognition [26]. In order to avoid the matrix inverse operation in CRC and to lead to more numerically stable solution, we proposed iterative CRC. Based on the idea to minimize the representation residual and the gradient descent algorithm, we know that the formula to iteratively calculate the coefficient of CRC is as follows.

First we use the Eq.(6) to calculate the coefficient Ψ as the original coefficient of ICRC. The error vector reconstructive representation can be calculated as follows:

$$e = y - \Psi X \quad (9)$$

Then we can calculate coefficient variation according to the error vector e :

$$\Delta \Psi = X^T e \quad (10)$$

At last we can update the coefficient:

$$\hat{\Psi} = \Psi + \lambda \Delta \Psi \quad (11)$$

Where λ is a positive scalar number that balances the learning efficiency and convergence. In our experiments, we set $\lambda = \frac{1}{i}$, which i is the number of iterations. The learning efficiency of ICRC will change with the increase of iteration number. So ICRC has better convergence properties and is numerically stable.

As we know, the solution obtained using the iterative method is influenced by the initialization step. Moreover, though the solution obtained using the original CRC algorithm has its drawback, it is still somewhat reasonable. Thus, our iterative method takes the solution of the original CRC algorithm as the initialization to iteratively calculate the ultimate solution. As our iterative method starts from an acceptable initialization, it is easier to reach the optimal solution,

The iterative algorithm can be described as follows:

```

Task: Renew the Coefficient  $\psi$ 
 $\psi_0 = \psi$           ( $\psi_0$  is the original coefficient)
for i=1:iteration
    if i>1
         $\hat{y} = \psi \times X$           (Reconstruct the testing sample)
         $error = y - \hat{y} / \text{norm}(\hat{y})$   (Calculate the reconstructive error.  $\text{norm}()$  is the normalize function)
         $y = \hat{y}$ 
    else
         $[col, row] = \text{size}(X)$   (Get the column and row of X)
         $error = \text{zeros}(col \times row, 1)$  (If i=1,  $\psi$  renews nothing)
    end
     $\psi = \psi + X' \times error / i$       (Renew the coefficient)
     $\psi = \psi / \text{norm}(\psi)$           (Normalize the new coefficient)
    If  $\text{norm}(\psi_0 - \psi) \leq 0.005$ 
        Break
    end
     $\psi_0 = \psi$ 
end

```

FIG.1 THE CORE OF OPERATIONS OF ITERATIVE CRC ALGORITHM

From Fig. 1 we can see that we do not calculate the inverse matrix to update the coefficients. So ICRC can reduce the possibility of getting the unstable numerical computation. Normalization of $\hat{y}$ and ψ are the important steps of ICRC. These steps ensure that the algorithm gets the reasonable error. In other word, the normalization steps make y and $\hat{y}$ in $error = y - \hat{y} / \text{norm}(\hat{y})$ are almost in the same range of ℓ_1 -norm. When $i > 1$ the algorithm calculates the reconstructive error first, and then projects the training samples on the reconstructive error to get the temp coefficient. At last the algorithm updates coefficients to complete an iterative process. Due to the normalization, the reconstructive error does not reduce with the increase of iteration. So in the update process, the algorithm adds the weighted coefficient according to the iteration. When $i = 1$, the coefficients will update nothing. At this time ICRC equates to CRC.

The distance measure and the class identity can be gotten by Eq.7 and Eq.8 separately.

Experiments and Discussion

In this section, ICRC is evaluated on the two publicly available face datasets, AR and FERET face dataset. All face images from the two databases are properly aligned and cropped with only facial region retained. In order to illustrate the effectiveness of ICRC, we also compare with two state-of-arts: LRC and CRC.

Experiments on AR Database

The AR [27] dataset consists of over 4000 frontal images for 126 subjects. In our experiments, we choose a subset including 2400 non-occluded frontal views from 120 subjects. All the 20 images of a sample gallery subject are illustrated in Fig. 2.

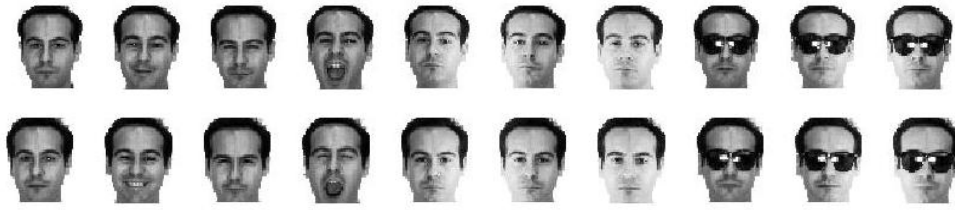


FIG. 2 SAMPLE IMAGES OF ONE SUBJECT IN AR DATASET.

TABLE 1 THE RECOGNITION RATE OF COMPETING METHODS ON AR DATABASE.

Method	$p=2$	$p=3$	$p=4$	$p=5$
LRC	0.5924	0.5909	0.5811	0.5956
CRC	0.7708	0.7642	0.7594	0.7667
ICRC	0.7931	0.7975	0.7859	0.7800

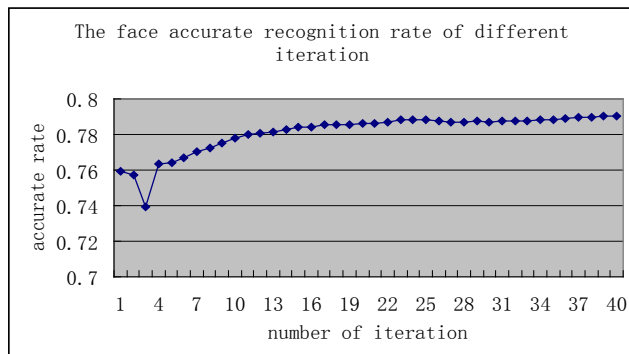


FIG. 3 THE RECOGNITION RATE OF DIFFERENT ITERATION (FROM 1 TO 40). THE TRAINING SET INCLUDES THE FIRST 4 FACE IMAGES OF EVERY CLASS ON AR FACE DATABASE.

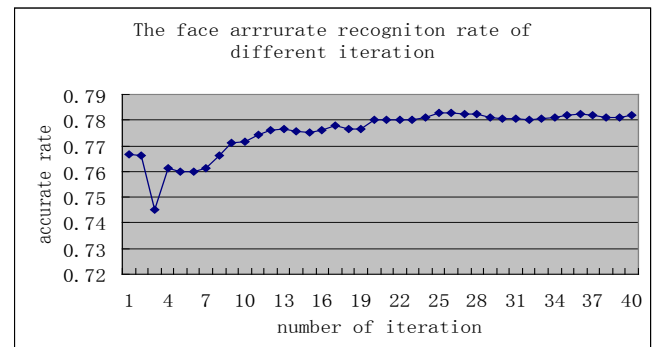


FIG. 4 THE RECOGNITION RATE OF DIFFERENT ITERATION (FROM 1 TO 40). THE TRAINING SET INCLUDES THE FIRST 5 FACE IMAGES OF EVERY CLASS ON AR FACE DATABASE.

In this experiment, we use the first p images of each class to form the training set, and the remained images of each class to form the validating set. The number of iteration is set as 20. The accuracy recognition rates are show in Tble 1.

We can see that CRC and ICRC are both at least 30% higher than LRC. The role of collaborative representation in the face recognition is clearly immense. As the improvement of CRC, the recognition rate of ICRC is higher than that of CRC at all events. In the best cases, the recognition rate of ICRC can be improved over 4%.

In order to observe and learn the relationship between different number of iteration and the accurate recognition rate, two experiments has been done and the results are shown as follows.

We take the first 4 and 5 face images of every class as the training set and take the others as testing set. The number of iteration is from 1 to 40, Fig 3 and Fig. 4 show the results of the two experiments.

Fig.4 and Fig.5 show that the accurate rate will reduce with the increase in the number of iterations when the number of iteration less than 4. Then the accurate recognition rate will increase along with the increase of iteration.

Experiments on ORL database

The ORL database contains 40 persons, each person including 10 different images [28]. The images of the same person are taken at different times, under slightly varying lighting conditions and with various facial expressions. Some persons are captured with or without glasses. The heads in images are slightly tilted or rotated. The images in the database are manually cropped. Fig. 4 shows ten images of a sample gallery subject.

In this experiment we take the last p images of each class as the training samples and the remained images for each person as the testing samples. The accuracy recognition rates are show in Table 2.

From the results of Table 2, we can see that the advantages of CRC can not be shown. When $p = 2$ and $p = 4$ the recognition rate of CRC is even less than that of LRC. In other cases, the accuracy of SRC and LRC are the same.

But the recognition rate of ICRC is higher than that of LRC and CRC except in the case of $p = 2$.

From Fig. 4 and Fig. 5 we can learn that experiments on ORL face database have the similar law with experiments on AR face database. But ICRC on ORL face database, the accuracy recognition rate stabilizes rapidly and gets the top rate. Compared with ORL face database, AR face database have more large number of classes and samples. This may be the major reason why the results of Fig.2 and Fig.3 are less stable than that of ORL face database.



FIG. 5 SAMPLE IMAGES OF ONE SUBJECT IN THE ORL DATASET.

TABLE 2 THE RECOGNITION RATE OF COMPETING METHODS ON ORL DATABASE.

Method	$p=2$	$p=3$	$p=4$	$p=5$
LRC	0.7750	0.8571	0.9292	0.9300
CRC	0.7656	0.8571	0.9250	0.9400
ICRC	0.7656	0.8929	0.9458	0.9500

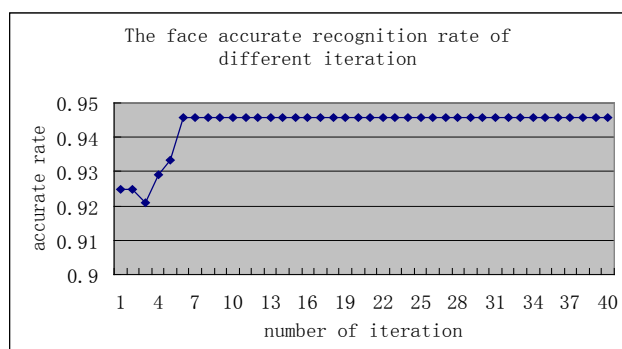


FIG. 4 THE RECOGNITION RATE OF DIFFERENT ITERATION (FROM 1 TO 40). THE TRAINING SET INCLUDES THE LAST 4 FACE IMAGES OF EVERY CLASS ON ORL FACE DATABASE.

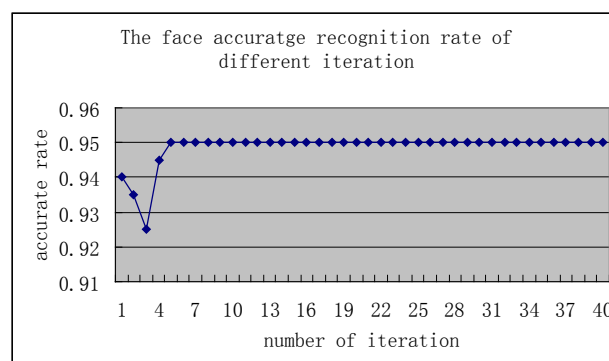


FIG. 5 THE RECOGNITION RATE OF DIFFERENT ITERATION (FROM 1 TO 40). THE TRAINING SET INCLUDES THE LAST 5 FACE IMAGES OF EVERY CLASS ON ORL FACE DATABASE.

Conclusions

In this paper we propose ICRC for face recognition. The experimental results show that ICRC obtains more accurate recognition results than CRC and LRC in face recognition. This shows that in ICRC the increase of the numerical stability also contributes to the improvement of recognition results. In addition, it seems that the performance of ICRC may be further improved and we will study it in the future.

ACKNOWLEDGMENT

This work is partially supported Specialty Construction Project in Guangdong Province (No.2013LYM0055 and 2013KJCX0127), Scientific Project of Hanshan Normal University (LY201301), the Science and Technology Program of Chaozhou City (2012S13) and Natural Science Foundation of Guangdong Province (No. 2014A030307014, 2014A030310038).

REFERENCES

- [1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips. "Face recognition: a literature survey." ACM Computing Surveys, 35(4), 399-458 (2003).
- [2] Y. Xu, X. Zhu, Z. Li, G. Liu, Y. Lu, H. Liu. "Using the original and "symmetrical face" training samples to perform representation based two-step face recognition." Pattern Recognition, 2013(46),1151-1158.
- [3] M. Kan, J. Wu, S. Shan, X. Chen. "Domain Adaptation for Face Recognition: Targetize Source Domain Bridged by

- Common Subspace." *International Journal of Computer Vision*, 109(1-2), 94-109 (2014).
- [4] J. Yang, D. Chu, L. Zhang, Y. Xu, J.-Y. Yang. "Sparse Representation Classifier Steered Discriminative Projection With Applications to Face Recognition." *IEEE Transactions on Neural Networks and Learning Systems*, 24(7), 1023-1035, 2013.
 - [5] R. R. Jillela, A. Ross. "Daptive frame selection for improved face recognition in low-resolution videos." *IJCNN*, 2009: 1439-1445.
 - [6] Y. Xu, X. Li, J. Yang, Z. Lai, D. Zhang. "Integrating conventional and inverse representation for face recognition." *IEEE Transactions on Cybernetics*, 44(10), 1738-1746 (2014).
 - [7] S. Wang, H.-L.Chen, W.-J. Yan, Y.-H. Chen, X. Fu. "Face Recognition and Micro-expression Recognition Based on Discriminant Tensor Subspace Analysis Plus Extreme Learning Machine." *Neural Processing Letters*, 39(1), 25-43 (2014).
 - [8] X. Zhang, Y. Gao. "Face recognition across pose: A review." *Pattern Recognition*, 42(11), 2876-2896 (2009).
 - [9] M. Turk, A. Pentland. "Eigenfaces for Recognition." *Journal of Cognitive Neuroscience*, 3(1), 71-86 (1991).
 - [10] P. Comon. "Independent component analysis: a new concept." *Signal Processing*, 36(3) 287-314 (1994).
 - [11] D.L. Setes, J. Weng. "Using discriminant eigenfeatures for image retrieval." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), 831-836 (1996).
 - [12] Z. Jin, J.Y. Yang, Z.S. Hu, Z. Lou. "Face recognition based on uncorrelated discriminant transformation." *Pattern Recognition*, 34(7), 1405-1416 (2001).
 - [13] Y. Xu, X. Fang, Q. Zhu, Y. Chen, J. You, H. Liu. "Modified minimum squared error algorithm for robust classification and face recognition experiments." *Neurocomputing*, 135(5), 253-261 (2014).
 - [14] F. Shen, O. Hasegawa. "A fast nearest neighbor classifier based on self-organizing incremental neural network." *IEEE Transactions on Neural Networks*, 21(10), 1537-1547 (2008).
 - [15] S.Z. Li, J. Lu. "Face recognition using the nearest feature line method." *IEEE Transactions on Neural Networks*, 10(2), 439-443 (1999).
 - [16] Y. Mitani, Y. Hamamoto. "A local mean-based nonparametric classifier." *Pattern Recognition Letters*, 27(10), 1151-1159 (2006).
 - [17] Z. Lou, Z. Jin. "Novel adaptive nearest neighbour classifiers based on hit-distance." In: *Proceedings of the 18th International Conference on Pattern Recognition*, 3, 87-90 (2006).
 - [18] Y. Xu, Q. Zhu, Z. Fan, M. Qiu, Y. Chen, H. Liu. "Coarse to fine K nearest neighbor classifier." *Pattern Recognition Letters*, 34(9), 980-986 (2013).
 - [19] Q. Gao, Z. Wang. "Center-based nearest neighbor classifier." *Pattern Recognition*, 40(1) 346-349 (2007).
 - [20] R. Togneri. "Linear Regression for Face Recognition." *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(11), 2166-2121 (2010).
 - [21] J. Wright, Y. Yang, S.S. Sastry, Y. Ma. "Robust face recognition via sparse representation." *IEEE Trans. Pattern Analysis and Machine Intelligence*, 3(2), 210-227 (2009).
 - [22] L. Zhang, M. Yang, X. Feng. "Sparse representation or collaborative representation: which helps face recognition?" In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 471-478 (2011).
 - [23] Y. Xu, Q. Zhu, Z. Fan, D. Zhang, J. Mi, Z. Lai. "Using the idea of the sparse representation to perform coarse to fine face recognition." *Information Sciences*, 238(20), 138-148 (2013).
 - [24] J. Wang, K. N. Plataniotis, J. Lu, A. N. Venetsanopoulos. "Kernel quadratic discriminant analysis for small sample size problem." *Pattern Recognition*, 41(5), 1528-1538 (2008).
 - [25] M. Aharon, M. Elad, A. Bruckstein. "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation." *IEEE Transactions on signal processing*, 54(11), 1311-4320 (2003).
 - [26] Y. Xu, D. Zhang, J. Yang, J. Y. Yang. "A two-phase test sample sparse representation method for use with face recognition." *IEEE Transactions on Circuits and Systems for Video Technology*, 21(9), 1255-1262 (2011).

[27] http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html

[28] <http://www.cam-orl.co.uk>

Wei Huang is now an associate professor with the department of computer science and engineering, Hanshan Normal University, China. He received his M.S. degree in computer application technology in 2008 from Nanjing University of Science and Technology, and his B.Sc. degree in Hanshan Normal University in 2001. His research interest includes pattern recognition, machine learning and image processing.

Xiaohui Wang received her B.S. degrees in mechanics engineering from the Hunan University, Changsha, China, in 2000, and the M.S. Degrees in Computer Science from the Harbin institute of technology, Harbin, China, in 2006. From 2000 to 2003, she was an assistant at the department of Physics in the south of Fujian Normal University. From 2006 to now, she served as lecturer at institute of computer and information engineering in Hanshan Normal University. Her research interests include image processing, pattern recognition and statistical signal processing.

Yinghui Zhu received her B.S. and M.S. degrees in computer science from South China Normal University, Guangzhou, China, in 1999 and 2005 respectively. He is now an associate professor of the department of computer science and engineering of Hanshan Normal University, China. His research interests include graphic and image processing, information security.

Jianzhong Li received the B.S. degree in Computer Science from Shantou University in 2005, and the Ph.D. degree in Optics from the Xiamen University, China, in 2011. He is now an associate professor of the department of mathematics and statistics of Hanshan Normal University, China. His research interests include image processing, information security and pattern recognition.